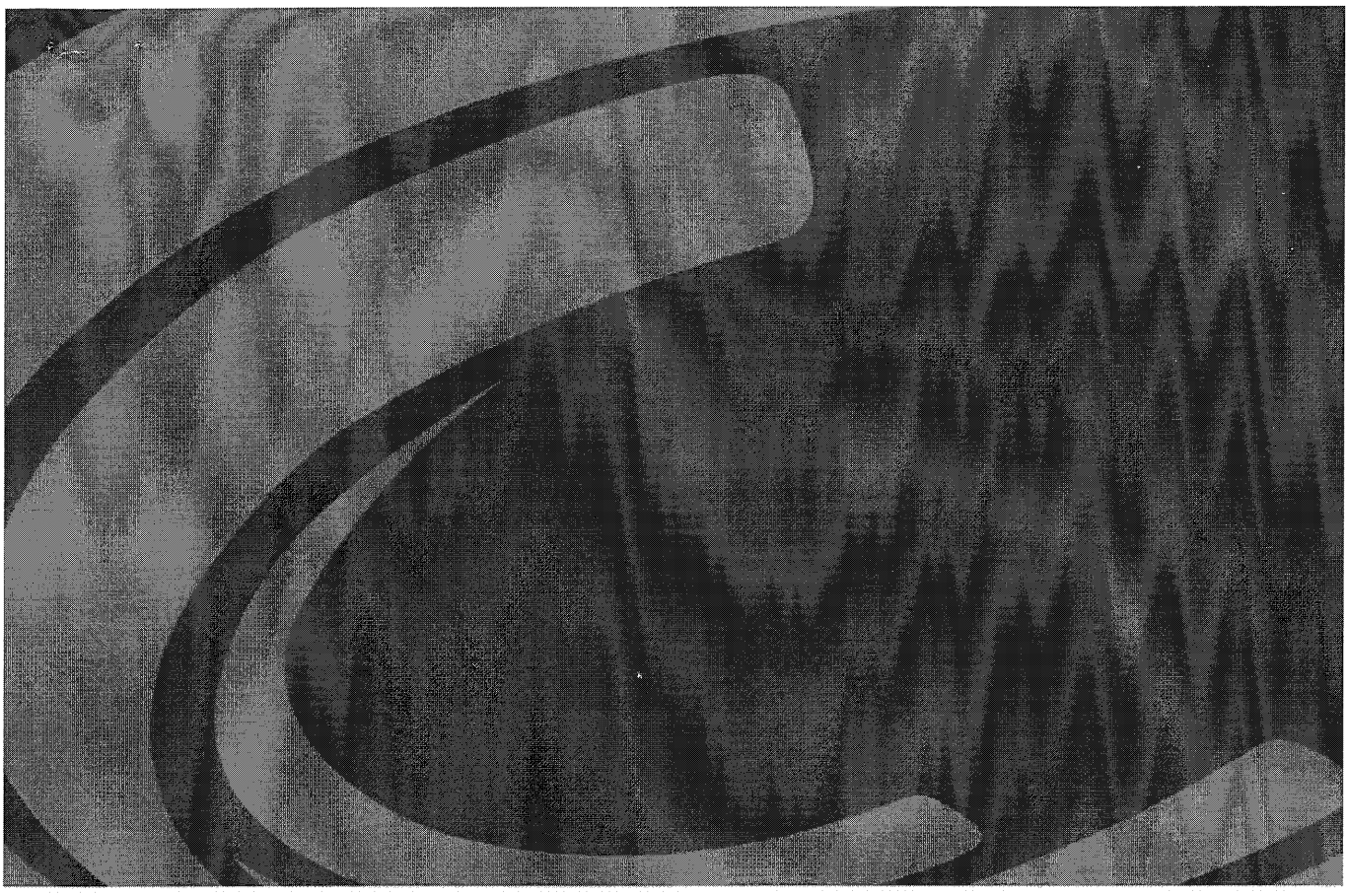
# EXHIBIT 16



# U.S. PREVENTIVE SERVICES TASK FORCE

## PROCEDURE MANUAL

U.S. Preventive Services
TASK FORCE

December 2015

# Foreword

Created in 1984, the U.S. Preventive Services Task Force (Task Force) is an independent, volunteer panel of 16 national experts in prevention and evidence-based medicine. Our mission is to improve the health of all Americans by making evidence-based recommendations on clinical preventive services and health promotion in primary care settings.

The Task Force is committed to making the recommendation development process as clear and transparent as possible so that health care professionals, partners, and the American public are fully informed every step of the way. We share the USPSTF Procedure Manual with this goal in mind.

This Procedure Manual describes the methods used by the Task Force to ensure that its recommendations are scientifically sound, reproducible, and well documented. It is intended as a guide for anyone who is interested in the Task Force, Task Force members, and those who support the Task Force's work, including staff of the Agency for Healthcare Research and Quality and its designated Evidence-based Practice Centers.

The Manual provides a high-level description of the Task Force's structure, governance, and processes for selecting topics, reviewing evidence, soliciting and responding to public input, and arriving at a recommendation. Researchers seeking a more detailed description of methods used to conduct a systematic evidence review may want to review the methods described on the Web site of AHRQ's Effective Health Care Program (effectivehealthcare.ahrq.gov) or read the evidence reviews that are posted with each final Task Force recommendation.

It is important to the Task Force that our colleagues, partners, and the American public understand our procedures. We hope that you will find the USPSTF Procedure Manual helpful and will share it with others who may find it beneficial. If you have any questions or comments, please contact the USPSTF Coordinator at coordinator@uspstf.net.

Together, we can work to improve the health of all Americans.

Albert L. Siu, M.D., M.S.P.H.,
Chair, U.S. Preventive Services Task Force

Michael LeFevre, M.D., M.S.P.H.
Immediate Past Chair, U.S. Preventive Services Task Force

For preventive interventions, the population benefit may be further limited by such issues as the following:

1. The prevalence and incidence of the target condition
2. For heterogeneous conditions, the prevalence of the condition subtype that would cause important health problems
3. The sensitivity of the screening test (i.e., the degree to which the test or a given threshold to define abnormality of the screening test will detect the subtype of the condition that would potentially cause health problems; sensitivity is rarely 100%)
4. The comparative effectiveness of early treatment of asymptomatic disease relative to later treatment of symptomatic disease of the subtype of the condition that would cause health problems (rarely 100%)

#### 6.5.4 Conceptual Confidence Limits

As previously noted, estimates of magnitude of benefit are intrinsically more uncertain when direct evidence is limited or absent or restricted to select populations or clinical scenarios. In these cases, the Task Force may place conceptual upper or lower bounds on the magnitude of benefit as applied to the population targeted in the recommendation. Considerations such as baseline risk of study participants and the clinical setting in which the studies were conducted also factor into the bounds of estimates of magnitude of benefit. For example, if magnitude of benefit is estimated only from studies of an intervention conducted by highly trained clinicians using specialized equipment for persons at considerably increased risk, this estimate might be considered the upper bound for benefit that might reasonably be anticipated for a general population. In other situations, the Task Force may also logically judge the lower bounds of the benefit, particularly when estimating the anticipated benefits in a population with a lower prevalence of disease than the study population in which the estimate of the benefit was derived.

Screening for abdominal aortic aneurysm is an example of the Task Force's use of conceptual confidence intervals. The benefits observed in screening studies of male smokers that were conducted in academic centers with optimal diagnostic and surgical treatment capabilities were judged to likely represent the upper bounds of benefit if these services were to be provided more generally in community-based settings. A lower conceptual bound of potential benefit was judged when extrapolating these studies in a high-risk population (male smokers) to populations at lower risk (male nonsmokers and female smokers and nonsmokers).

#### 6.5.5 Outcomes Tables and Decision Modeling for Determining Magnitude of Benefit

One way to determine the magnitude of benefit is to use an outcomes table based on the systematic evidence review or, when available, outputs from a decision model. An outcomes table can demonstrate how many or the proportion of persons likely to benefit—and in what ways—from implementation of the preventive service. Estimates from direct and indirect evidence may be included in outcomes tables in order to provide the range of expected magnitude of specific beneficial outcomes (**Appendix VIII**).

### 6.6 Assessing Magnitude of Harm

#### 6.6.1 Definitions of Magnitude Ratings and Criteria

The Task Force starts with the conceptual notion that screening, counseling, or use of preventive medications are intended for asymptomatic individuals in order to prevent or delay future health problems. The burden of proof that the benefits exceed the harms prior to recommending implementation of screening or other preventive services is thus higher than it is for diagnosis or treatment of symptomatic conditions. As such, assessment of the magnitude of harm is critically important. As with the magnitude of benefit, in situations where the evidence is adequate or convincing for harm, the magnitude of harm is assessed using the following categories: substantial, moderate, small, or zero. If the evidence is deemed inadequate for the assessment, the magnitude of harm rating is not applicable.

The Task Force uses the evidence to estimate the size of the population that would be harmed from implementation of the preventive service over a given time horizon (appropriate to the service under consideration) and over the expected time to be harmed and the duration/severity of the harm. Assessment of the magnitude of harm may be more difficult than assessment of benefit for many reasons. The broad range of potential harms is often less well identified or reported than potential benefits. At times severe harms occur at a relatively infrequent rate compared to benefits and require larger sample sizes than those studied in RCTs designed to evaluate benefits. Unlike fairly discrete benefits that the preventive service is intended to provide, harms are often varied and complex, occur at several stages in the screening cascade (including at earlier times than for benefits), may persist, and may be poorly recognized. Furthermore, for many, understanding that screening and preventive tests and procedures can cause harm is conceptually difficult.

As with benefit, the magnitude of harm might be determined directly from the reported results of large well-conducted RCTs of a preventive intervention, but more often also requires an assessment across the key questions and the linkages in the analytic framework (even when RCTs are available). Nonrandomized studies are often considered a more reliable source of detecting and determining the magnitude of harm (especially rare but serious harms) than for assessment of benefit. Data on harms may be inadequate for an assessment of magnitude, even when there is adequate data to characterize benefit because of the variability in the reporting of harms and the fact that many studies are not statistically powered or designed to detect some harms.

#### 6.6.2 General Types of Harm for Consideration

The Task Force starts with the assumption that nearly all preventive interventions have the potential to result in some magnitude of one or more harms to patients. For screening-based recommendations, the Task Force looks for harms of the screening test, the subsequent diagnostic tests resulting from screening, and early treatment of screen-detected asymptomatic disease. For recommendations that involve preventive medications and behavioral interventions, the Task Force looks at the magnitude of harm from these interventions.

Harms of screening may include psychological harm from labeling, the harms of diagnostic studies to confirm the presence of the condition, and overdiagnosis of screen-detected conditions. Because screening and other preventive interventions are implemented in asymptomatic persons with the goal of preventing future disease, the Task Force places a high priority on the effects of overdiagnosis and overtreatment, whereby the preventive service has the unintended consequence of creating "disease" that often leads to unnecessary and ineffective treatment. Harms of early treatment and overdiagnosis may accrue to patients whose condition might never have come to clinical attention or for whom the harms of treatment initiated prior to routine clinical detection were different or occurred earlier and/or over a longer period of time. In other words, these are harms of treatment that would not have occurred in the absence of screening.

Harms may also be considered in the form of opportunity costs for both patients and providers. The Task Force may consider the time and effort required by both patients and the health care system to implement the preventive care service. If the time and effort are judged to be substantially greater than other preventive services delivered in the primary care setting, these factors are also considered in the harms category. The Task Force usually derives qualitative, rather than precise, estimates of opportunity costs.

Although opportunity costs may be considered in the determination of Task Force recommendation grades, financial costs are not. Financial costs are also not considered in the decision models used for Task Force recommendations.

#### 6.6.3 Conceptual Confidence Intervals in Face of Inadequate Direct Evidence of Harms

Although there is often less evidence about potential harms than about potential benefits, the Task Force may draw general conclusions from evidence on expected yield of screening in terms of false-positive test results. If the prevalence of the condition is low and the specificity of the test is less than 100%, the positive predictive value may be low and false-positive test results will be expected. If the diagnostic workup is invasive or otherwise carries clinically important potential for harm, the Task Force can infer that at least some harms will result from implementation of the screening program, because some persons with false-positive screening tests will undergo an invasive diagnostic protocol for no possible benefit.

Similarly, if overdiagnosis (and therefore overtreatment) is common, and if the treatment has some adverse effects, the Task Force may infer that implementation of routine screening will cause at least some incremental harms, even in the absence of studies that characterize harms. This approach does not require an exact estimate of the magnitude of harm, but rather a determination that the harms are unlikely to be less than what is known about the number of false-positives, the invasiveness of the diagnostic workup, and the expected amount of overtreatment. Care should be taken to call attention to the estimate's lack of precision.

#### 6.6.4 Presentation of Harms in Outcomes Tables

As with the magnitude of benefit, the magnitude of harm may be informed by an outcomes table based on the systematic review or, when available, outputs from a decision model. When outcomes tables are used to present benefits, estimates for harms will also always be presented.

### 6.7 Assessing Certainty of Evidence for the Entire Analytic Framework

#### 6.7.1 Overview

The Task Force defines *certainty* as "likelihood that the USPSTF assessment of the net benefit of a preventive service is correct." The *net benefit* is defined as the benefits minus the harms of the preventive service as